Autonomous spatially-adaptive sampling in experiments based on curvature, statistical error and sample spacing with applications in LDA measurements

Raf Theunissen, Jesse S. Kadosh, Christian B. Allen

Department of Aerospace Engineering, University of Bristol, Bristol, UK
* Correspondent author: r.theunissen@bristol.ac.uk

Abstract
Spatially varying signals are typically sampled by collecting uniformly spaced samples irrespective of the signal content. For signals with inhomogeneous information content, this leads to unnecessarily dense sampling in regions of low interest or insufficient sample density at important features, or both. A new adaptive sampling technique is presented directing sample collection in proportion to local information content, capturing adequately the short-period features while sparsely sampling less dynamic regions. The proposed method incorporates a data-adapted sampling strategy on the basis of signal curvature, sample space-filling, variable experimental uncertainty and iterative improvement. Numerical assessment has indicated a reduction in number of samples required to achieve a predefined uncertainty level overall while improving local accuracy for important features. The potential of the proposed method has been further demonstrated on the basis of Laser Doppler Anemometry experiments examining the wake behind a NACA0012 airfol and the boundary layer characterisation of a flat plate.

Keywords: Adaptive sampling, curvature, uncertainty, radial basis function, LDA, boundary layer, airfoil

1 Introduction
Independent of the adopted measurement technique, data extraction is a discrete process by nature, i.e. data can only be captured at distinct temporal or spatial instances with finite resolution. The most common sampling routines adopt so-called space filing strategies whereby sampling locations are attributed throughout the measurement domain irrespective of the underlying signal. Such strategies include full-factorial i.e. equispaced locations in each dimension, stratified, Latin Hypercube designs (McKay *et al.* 1979), Sobol sequences or fully random sequences. These sampling approaches are typically not satisfactory or not practical though because of the exponentially large number of samples required to properly characterise the signal. Iterative methods on the other hand, whereby extraction locations are determined on the basis of the accrued signal content itself, offer the potential to optimise the attribution of the number and position of sampling points. Accumulating information in regions of increased signal complexity offers a more efficient strategy and consequently minimises testing time without loss of information. Adaptive sampling is commonly adopted in clinical research to ensure appropriate sample sizes for efficient and meaningful statistical experiments (Chow and Chang 2008; Chow 2014). The efficient collection of meaningful data has however a broad application and is commonly referred to as Design of Experiments (DoE) for which vast literature is available (Weissman and Anderson 2014; Ilzarbe *et al.* 2008). For instance, in precision manufacture of parts, adaptive sampling routines have been developed to enhance the automated geometry inspection with coordinate measuring machines (Yu *et al.* 2013), while in chemical physics adaptive methods have been used to refine surrogate models for potential energy surfaces by sampling regions

of high wave energy content (Sparta *et al.* 2009). The authors will present within the current study the concept of such an adaptive routine to guide the spatial sampling of the velocity field in fluid dynamics experiments.

Independent of the application, all adaptive sampling methodologies inherently require first a metamodel or surrogate model of the signal of interest, followed by an indicator of complexity. Both are updated iteratively with newly collected data. Surrogate models are built by testing sample points of the target response, i.e. the signal, to provide a heuristic over the whole domain, albeit subject to uncertainty using all information available. The objective is to ensure new sample points are placed where they will be most advantageous in terms of new information content without the necessity to traverse the response landscape at infinitely many points. Surrogate modelling has received considerable attention in the field of engineering as it allows optimisation of a design while offering the benefit of being cheap to evaluate as a direct result of the reduction in required data points (Forrester *et al.* 2008). The use of Kriging as a particular metamodel is most common as the collected data is typically assumed to be a realization of Gaussian processes, i.e. each value of the underlying signal is assumed to follow a normal probability density (Picheny *et al.* 2010). Such methods however require further definition of the related covariance functions through the so-called hyperparameters (Toal *et al.* 2008). Alternatively one can resort to Radial Basis Functions (Fasshauer 2007) in combination with polynomial interpolation (Fox 1997). In comparison with other interpolation schemes such as splines, polynomial response models and Kriging, Radial Basis Functions (RBF) offer the advantage of being conceptually simple in construction, meshless (data can be arbitrarily spaced), easily extendible to higher dimensions and provide models of arbitrary smoothness (Wendland 2005). Developed first to reconstruct complex geographical landscapes (Hardy 1971), RBFs typically out-perform polynomials in terms of reconstruction accuracy (Hussain *et al.* 2002); and even the ability of polynomials to model simple, convex responses of low dimensionality (Hussain *et al.* 2002; Paiva 2009) can be subsumed into the RBF approach.

With a surrogate model at hand, the localisation of the samples is next dictated by the objective function which prioritises new potential sampling positions. Lovison and Rigoni (2010) for example apply the Lipschitz criterion as indicator for the signal's complexity whereby more points are allocated to regions of higher complexity. In PIV image processing Theunissen *et al.* (2007) attributed higher sampling densities to areas of higher variance in velocity and achieved higher spatial resolution. However, adaptivity criteria were inherently defined by first order gradients while discretization requires denser sampling in regions of stronger curvature due to nonlinearity of the signal. Mackman and Allen (2010), consult the Laplacian to guide local refinement and model the spacing between samples with an interpolant. While such methods offer drastic improvements in the case of deterministic data, the introduction of experimental noise presents special challenges as random fluctuations contaminate correct measures of spatial curvature, and as regions of high and low uncertainty are treated without account of their differences.

The objective of the work presented is to address the spatial allocation of velocity extraction through an automated adaptive process incorporating the mathematical formulations of typical decision criteria applied by the experimentalist. Curvature of the velocity field is used to adjust the budget and spatial organisation of data extraction sites. The approach presented herein further alleviates constraints on sampling imposed by spurious curvature measures by including the local measurement error as a heuristic driving the sampling distribution. Moreover, the formulation of the error objective used here admits heteroskedastic error profiles: rather than considering only a domain-wide scalar value for uncertainty, locally varying error information can be exploited to focus more samples in regions of higher uncertainty or volatility, such as regions of unsteady flow. In an iterative manner the number of attributed samples is incremented using a Radial Basis Function interpolant as artificial target signal from which the adaptivity criteria are derived. To the best of the authors' knowledge,

adaptive measurements utilising error estimates have not yet been applied in the field of experimental fluid dynamics.

To clarify the framework of the presented work, the sampling problematic is presented in section 2. Details regarding the implementation of the surrogate model are provided in section 3. This section contains the novel incorporation of measurement error into the objective function as well as an iterative convergence criterion to automate the sampling procedure. The conduciveness of the presented methodology is subsequently assessed on the basis of computer simulated signals and one-dimensional LDA experiments behind a NACA airfoil and flat plate boundary layer showing the achievable gain in spatial resolution while simultaneously minimising testing time.

2. Problem statement

In the engineering community, the Nyquist sampling criterion is well-known and dictates the required distance between neighbouring sample points on the basis of the maximum occurring signal frequency (Shannon, 1949). In order to capture the features of a sinusoid of frequency B in a domain of length L, for example, the minimum required sampling rate is $2B$. This represents an ideal case with perfect placement of all samples in a uniform manner with spacing $L/2B$. In case of inhomogeneous distribution of information, e.g. a signal with spatially varying frequency, such an approach will however lead to inadequate sampling as the imposed sampling criterion will be too stringent in signal portions of lower frequency or lower interest. The experimentalist is *a-priori* uninformed of the signal's spectral characteristics and will concomitantly adopt an over-sampling strategy considered safe, though inherently severely misusing computational and acquisition effort. More importantly, sampling abiding the Nyquist criterion does not guarantee the reconstruction to provide a proper representation of the underlying flow field as illustrated in Figure 1-a. Sampling must concomitantly be also adapted to the signal complexity, i.e. data dependent sampling.

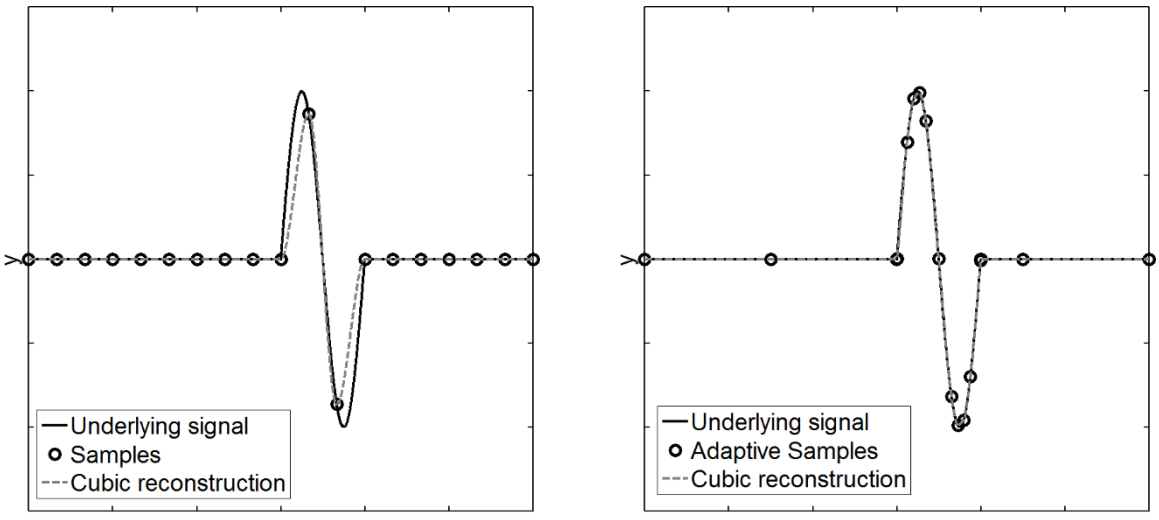


Figure 1: (a) Example of a signal containing spatially inhomogeneous data content, sampled at three times the signal frequency using full factorial sampling and reconstructed using cubic interpolation. **(b)** Adaptively sampled and reconstructed signal.

At this stage the authors would like to stress that because of the inherent sequential sampling, in the current framework data extraction focusses on the extraction of statistical quantities such as e.g. average velocity field. Only when an instantaneous full-field sample of the underlying signal is available, such as in PIV, can the sampling strategy yield optimised measurement sites for

instantaneous data (Theunissen *et al.*, 2007). Sampling concepts presented hereafter will nevertheless remain valid.

Adaptive sampling considers response data in the determination of new sample locations as to retain the level of accuracy with fewer required samples, or to produce a model with superior accuracy given the same budget of samples (Sacks *et al.*, 1989). These techniques add samples iteratively at locations where based on values from the current surrogate model a sampling criterion is met. The surrogate model is then updated with the new samples and responses for the next iteration. The outcome of such a procedure is depicted in Figure 1-b. Extraction locations are clustered in signal portions of sufficient interest yielding a suitable basis for reliable signal reconstruction.

To further exemplify the proposed dependency between sampling and signal content a one-dimensional parabola is considered. The parabola is defined as

$$f(x)=a(x-x_p)^2+b \quad (1)$$

where x_p symbolises the peak location. Sampling is performed at locations x_i allowing the underlying signal to be approximated as

$$g(x)=\Sigma f(x_i)\psi(x-x_i) \quad (2)$$

This formulation is the general descriptor of interpolation with $\psi(\cdot)$ representing a basis-function. Imposing the property $\Sigma\psi(x-x_i)=1$, which is valid for polynomial-based interpolation kernels, the error ε between the true and reconstructed value at location x is then given as

$$\varepsilon^2(x)=\{g(x)-f(x)\}^2=a^2\{\sum_i[(x_i-x_p)^2-(x-x_p)^2]\cdot\psi(x-x_i)\}^2$$
$$\varepsilon^2(x_p)=a^2\{\sum_i(x_i-x_p)^2\cdot\psi(x_p-x_i)\}^2 \quad (3)$$

Assuming $\psi(\cdot)>0$ and to be monotonically decreasing, largest errors will be attained at the location with the smallest radius of curvature $(x=x_p)$, in which case the error will be proportional to the local curvature, determined by parameter a, and distance between the sample and peak location (x_p-x_i). This simplified example justifies the need to agglomerate sampling instances in regions of higher signal curvature (Dassi *et al.*, 2014) and improves reconstruction as demonstrated in Figure 1-b.

The authors also consider data uncertainty as a secondary criterion to steer the sampling. Adaptive sampling is a concept greatly explored for engineering design optimisation in the CFD community for which the uncertainty in response values is wholly due to modelling errors and numerical inaccuracies. The estimate of error so obtained is based purely on the confidence in the modelling parameters and their ability to predict the responses at untested sites based on proximity to existing samples (Forrester *et al.*, 2006). This response data may be presumed to be homoscedastic i.e., the response uncertainty level is constant throughout the domain. Framed in this way, the confidence in response prediction depends entirely upon the spacing of samples and in no way reflects the value or uncertainty of their responses. This is evident in practice, as positioning sample updates based on the Kriging estimate of error alone results in a completely global i.e., space-filling, sample distribution. In contrast, data derived from physical measurements is subject to real stochastic error (Theunissen *et al.*, 2008). In studies of fluid flows, uncertainty can arise additionally from unsteadiness as in the case of turbulence. Following Bendat and Piersol (1966) the relative measurement error $\varepsilon_{\bar{x}}$ and uncertainty in mean statistics $\bar{x}$ is related to the second moment by

$$\varepsilon_{\bar{x}}=\frac{Z_\alpha}{\sqrt{N_s}}\frac{\sigma_x}{\bar{x}}=\frac{Z_\alpha}{\sqrt{N_s}}\frac{\sqrt{\overline{x'x'}}}{\bar{x}}\sim\frac{Z_\alpha}{\sqrt{N_s}}\frac{\sqrt{-\tau_{xx}}}{\bar{x}} \quad (4)$$

where N_s represents the number of independent samples at a fixed spatial location, Z_α a coefficient related to the confidence level (commonly set to 1.96 for a 95% confidence level)and σ_x the standard deviation in the data. Omitting the density factor the magnitude of turbulence intensity is linked to

Reynolds stress τ_{xx}. The latter may vary spatially and can thus be heteroskedastic. In these cases the authors argue that the assessment of error when deciding where to interrogate the physical system for new samples is sensible. Moreover, it is good practice to re-acquire data points in areas of elevated uncertainty to improve confidence in the data.

The budget of samples is typically dictated by the user, implying that significant knowledge of at least the form of the response is available before adaptive sampling begins. The number of samples is consequently rarely optimal. A last requirement of any automated adaptive routine, in the interest of maximum generality and making the method relevant to as wide a field of application as possible, is therefore a means of automatically determining when enough samples have been collected.

3. Proposed Methodology

The above discussion highlights the need to suppress the common constraint of data-independent sampling. For this reason an adaptive sampling strategy is proposed taking into account local curvature and spatially varying error estimations while avoiding unnecessary refinement (over-sampling). Initial sampling sites are distributed uniformly and define the extents of the domain. The task of the initial grid is to provide crude insight into the response with a minimum of samples but with a fair chance of capturing most important features. When the average domain coverage per sampling point reduces to 1.6%, the initial grid surpasses the accuracy requirement, negating the need for adaptive sampling. For this reason typical initial grid sizes of 17 to 33 sampling sites are considered hereafter. The subsequent adaptive sampling is driven by an objective function. Given the discrete availability of data, a surrogate model is needed to enable evaluations throughout the scrutinised domain. Because the model will require updating after the acquisition of data at new sampling points, the adaptive data capturing process is implemented in a recursive scheme as presented in Figure 2.

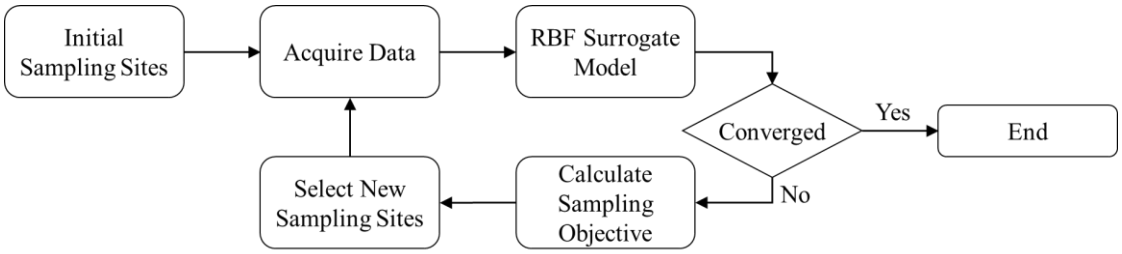


Figure 2: Proposed adaptive data acquisition using RBF surrogate modelling.

The aim of an adaptive sampling method with broad applicability demands that case-specific parameter tuning of the metamodel be avoided in favour of robust, universally relevant constants. The choice of surrogate model and adopted methodology of blending sampling criteria into one objective will be discussed hereafter in addition to an automated stopping criterion.

3.1 RBF Surrogate model

Curve fitting in the form of polynomial interpolation or spline fitting is probably the most ubiquitous model due to simplicity and relatively low cost. However, Radial Basis Functions (RBF)-based surrogate models are able to capture very complicated higher-dimensional landscapes out of reach for simpler models (Krishnamurthy 2002, 2005) while offering superior accuracy when combined with polynomial terms (Hussain *et al.*, 2005). Second, RBF and Kriging models offer single, domain-wide analytic expressions contrary to splines which are only piecewise analytic. A third benefit is that, as Gaussian process based models, RBFs provide credible error estimates. In fact, it can be shown that these error estimates are identical in form between RBF and Kriging models (Mackman *et al.*, 2013). In contrast to simpler models such as polynomial response surfaces, RBF and Kriging models do not

impose a particular functional form and are meshfree making them more conducive to approximate complex surfaces represented with scattered data.

If the field of application is particularly constrained it is possible to achieve very convincing predictions of the response by a-priori adjusting model parameters as in Kriging. The resulting model is tailored closely to a specific type of response however and the parameters so tuned are unsuitable for experiments with very different response shapes or scales. Congdon and Martin (2007) summarise the possible origins of deviations between the underlying response and a Kriging model response as being due to erroneous data, the form of the Kriging model being insufficient to estimate the observations as a Gaussian process, or the range of the model not being amenable to representation by a single spatial random process. To retain maximum generality and minimise assumptions about the data a priori, in the present work the authors resort to RBF as surrogate models.

For simplicity, the collection of N_p datapoints $x_{1..Np}$ are confined hereafter to a unit interval $x_s \in [0,1]$. This scaling is performed after every addition of data sites. The response $s(x_s)$ of a RBF interpolant combined with a polynomial p of a signal f, using N_p samples can be formulated as

$$s(x_s) = \sum_{k=1}^{N_p} \eta_k \psi(\| x_{k,s} - x_s \|) + p(x_s) \quad \text{with} \quad \sum_{k=1}^{N_p} \eta_k \, p(x_{k,s}) = 0 \qquad \textbf{(5)}$$

Here $\psi(||x_{k,s}-x_s||)$ is the contribution of the kth basis function depending solely on the radial distance (hence the naming) between the kth data point $x_{k,s}$ and new location x_s. As an interpolant $s(x_s)$ recovers the signal exactly at all data sites $x_{k,s}$ while a polynomial $p(x_s)$ is added to enhance accuracy. A third order polynomial $p(x_s)=q_3x_s^3+q_2x_s^2+q_1x_s+q_0$ is used hereafter. The interpolation problem presented in Equation (4) amounts to solving a system of linear equations

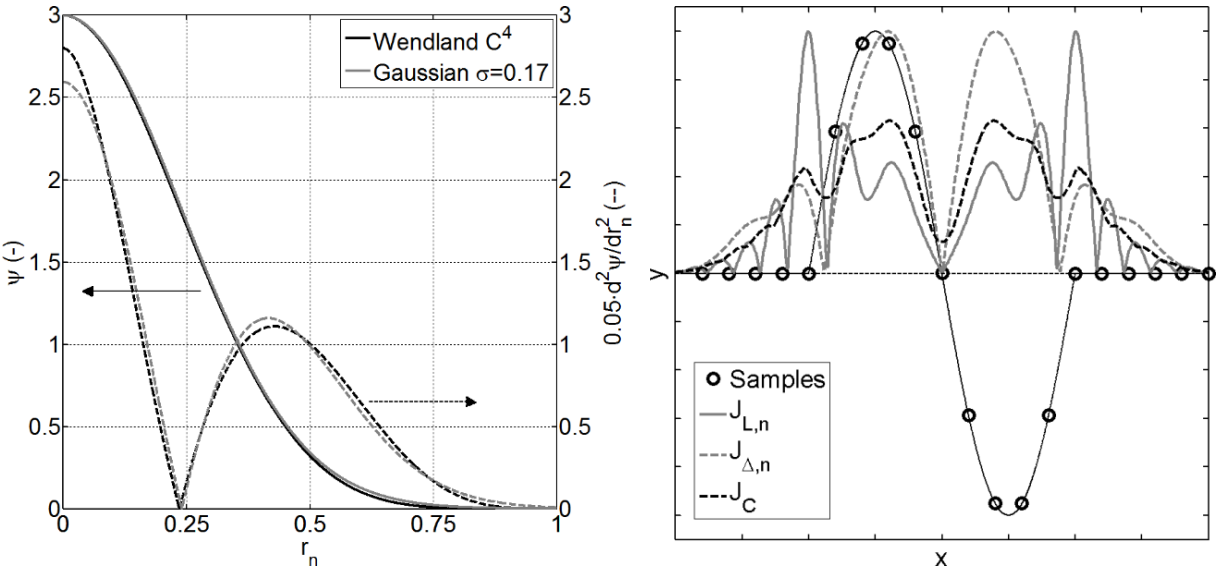
$$\begin{bmatrix} \Psi & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \underline{\eta} \\ \underline{q} \end{bmatrix} = \begin{bmatrix} \underline{f} \\ \underline{0} \end{bmatrix} \quad \text{equivalent to} \quad A_\psi \, \underline{m} = \underline{g} \qquad \textbf{(6)}$$

where Ψ is a $N_p \times N_p$ matrix with elements $\psi_{i,j}(||x_{i,s}-x_{j,s}||)$, P is a $N_p \times 4$ matrix with row-wise elements $[x_i^3 \; x_i^2 \; x_i \; 1]$ and $i,j=1 \ldots N_p$, O is a 4×4 zero matrix, vectors $\underline{\eta}=[\eta_1 \; \eta_2 \; \ldots \; \eta_{Np}]^T$, $\underline{q}=[q_3 \; q_2 \; q_1 \; q_0]^T$, $\underline{f}=[f_1 \; f_2 \; \ldots f_{Np}]^T$ and $\underline{0}$ is a vector containing four zero elements. The coefficients defining the interpolation can then be solved by means of a least squares approach; $\underline{m}=(A_\psi^T A_\psi)^{-1}(A_\psi^T \underline{g})$. After re-scaling the defined interrogation grid containing N_e potential sampling sites, $\underline{x}_{int}$, the surrogate model is evaluated as $s(\underline{x}_{int})=\Psi_{int}\underline{\eta}+P_{int}\underline{q}$ where the elements of the $N_e \times N_p$ matrix Ψ_{int} are defined as $\psi_{i,j}(||x_{int,i}-x_{j,k}||)$ and P_{int} contains the N_e row-wise elements $[x_{int,i}^3 \; x_{int,i}^2 \; x_{int,i} \; 1]$.

Available basis functions are many and varied, with popular choices including thin plate splines $\psi(r)=r^2 log(r)$, cubic $\psi(r)=r^3$, Gaussian $\psi(r)=exp(-r^2/2\sigma^2)$, multiquadric $\psi(r)=(r^2+\sigma^2)^{½}$, and others (Forrester *et al.* 2009; Hussain *et al.* 2002). In the current method, the fourth-order-continuous (C^4) radial basis function defined by Wendland (2005) is used in order to allow for a continuous and smooth analytic Laplacian which will serve as heuristic for the curvature. Wendland's basis functions are of minimal polynomial degree for a given smoothness. They are moreover compactly supported, ensuring the basis function is positive definite. This last property ensures solubility of the interpolation problem and also gives Wendland RBFs their relatively better matrix conditioning compared with other basis functions as the number of samples or the support radius are increased (Fornberg *et al.* 2002). Moreover, the availability of analytic derivatives of the RBF interpolant can be inexpensively computed without recalculating the interpolation coefficients. Interpolants are neither restricted to equidistant sample locations nor require any partitioning of the domain in a case-specific way. The implemented Wendland's fourth-order-continuous RBF for up to a maximum of three spatial dimensions is given by

$$\psi(r_n) = (1-r_n)_+^6 (35r_n^2 + 18r_n + 3) \quad \text{where} \quad (1-r_n)_+ = \begin{cases} 1-r_n & 1-r_n \geq 0 \\ 0 & 1-r_n < 0 \end{cases} \qquad (7)$$

Parameter r_n represents the normalised Euclidian distance. A comparison between the Wendland C^4 and Gaussian basis function with a standard deviation $\sigma=0.17$ is depicted in Figure 3-a for clarity.



(a) **(b)**

Figure 3: (a) Comparison between Wendland fourth-order continuous (C^4) basis function and a Gaussian with $\sigma=0.17$. Though both functions seem similar, the former has a finite non-zero extent up to $r_n=1$ while the Gaussian has an infinite support radius **(b)** Illustration of the blended curvature objective considering a windowed sinusoidal signal sampled at 10 times the sinusoid frequency.

The contribution from each basis site to the interpolant depends on the radial (Euclidian) distance scaled by a support radius R which determines the region of influence of each basis function;

$$r_n = \frac{\|x - x_i\|}{R} \qquad (8)$$

It has been shown that RBF interpolants have an optimal support radius value which minimises the interpolation error while retaining acceptable conditioning of the kernel matrix (Larsson and Fornberg 2005). This optimal value depends strongly upon the RBF kernel used, the distribution of data, the response values, and the number of data points though. Consequently, optimum selection of the support radius value has been a topic of ongoing discussion for many years with authors suggesting sophisticated parameter estimation schemes to suit each individual response (Rippa 1999; Sheuerer 2011) and others proposing to use a constant value that is robust over most responses (Mackman and Allen 2010). Setting a constant $R=1$ has been found previously to yield good results as each basis function can contribute throughout the entire domain. This also comes at the price of spurious Gibbs-like oscillations and poorer matrix conditioning. From the authors' experience $R=0.5$ is a robust choice for a wide variety of responses. This removes the expense of tuning R for each case or for each distribution of samples, and also strikes a balance between good accuracy throughout the domain and freedom from spurious oscillation phenomena and matrix conditioning problems.

3.2 Sampling Objective Function

In the current work individual objective factors J_i mathematically translate typical sampling criteria used by the experimentalist. These factors account for curvature (J_C), error (J_E), improvement (J_I) and

spacing criteria (J_h, J_{ho}) and are combined to yield the overall sampling objective function J on which to base the selection of new sampling sites;

$$J=(J_C+\kappa_C)\cdot(J_E+\kappa_E)\cdot(J_I+\kappa_I)\cdot J_h J_{ho} \quad (9)$$

where κ_i are empirical offset constants. The objective function J quantifies the significance of potential sampling sites across the domain; the higher the magnitude of J, the greater the benefit of sampling at the corresponding location.

- **Curvature adaptive sampling - J_C**

A first estimate of the heuristic for measuring curvature is given by the response's Laplacian;

$$J_L=|\nabla^2 s|=\left|\sum_{k=1}^{N_p}\eta_k\nabla^2\psi+\nabla^2 p\right| \quad (10)$$

In the current work considering the one-dimensional problem, p is the cubic polynomial yielding $\nabla^2 p=6q_3x_s+2q_2$. Once the interpolation coefficients η_k for the response have been found (section 3.1), calculation of the Laplacian proceeds in the same way as for the response whereby only the interrogation matrix needs to be recomputed using the derivative forms of the RBF basis functions. Symbolising the components of the N_d-dimensional coordinate vector as x_ξ with $\xi=1\ldots N_d$ and noting that $r^2=\sum_{\xi=1}^{N_d}x_\xi^2$ the Laplacian of the basis function is given by

$$\nabla^2\psi=\sum_{\xi=1}^{N_d}\frac{\partial^2\psi}{\partial x_\xi^2}=\sum_{\xi=1}^{N_d}\left[\frac{x_\xi^2}{r^2}\frac{d^2\psi}{dr^2}+\frac{r^2-x_\xi^2}{r^3}\frac{d\psi}{dr}\right]=\frac{d^2\psi}{dr^2}+\frac{N_d-1}{r}\frac{d\psi}{dr} \quad (11)$$

Although the Laplacian is the sum of second derivatives with respect to each coordinate direction, evaluation requires only first and second derivatives of the basis kernels with respect to separation distance. These derivatives are given analytically and for the Wendland C^4 basis function adopted in this work

$$\frac{d\psi}{dr}=-\frac{56}{R}r_n(5r_n+1)(1-r_n)_+^5 \quad \text{and} \quad \frac{d^2\psi}{dr^2}=\frac{56}{R^2}(35r_n^2-4r_n-1)(1-r_n)_+^4 \quad (12)$$

The second derivative is depicted in Figure 3-a and compared to that of a Gaussian for clarity. A smooth signal, such as might be obtained in deterministic (computer) experiments with no real noise, may have a smooth Laplacian, but physical measurements come with stochastic errors in the form of noise. These give rise to a model Laplacian which can be erratic and not always true to that of the underlying function. For this reason the new adaptive method blends the Laplacian with a measure J_Δ of local signal change relative to the wider signal environment;

$$J_\Delta=\frac{\Delta s(x_{\text{int}})}{|s_{MA}(x_{\text{int}})|+1}=\frac{|s(x_{\text{int}})-s_{MA}(x_{\text{int}})|}{|s_{MA}(x_{\text{int}})|+1} \quad (13)$$

s_{MA} represents the moving average filtered response evaluated on the interrogation grid considering a filter width equal to a fraction (~30%) of domain extent in the calculation of each point. The curvature objective function J_C is subsequently defined as

$$J_C=MA\left(\sqrt{J_{L,n}J_{\Delta,n}}\right) \quad (14)$$

Where the subscript n refers to unity-based normalisation, $(\cdot)_n=(\cdot)/(\cdot)_{max}$, and the square root avoids exacerbating relative differences in peak magnitudes. A secondary moving averaging filter MA with a span of 2% the domain has been implemented to improve tractability of the curvature for subsequent sample selection.

The concept is illustrated in Figure 3-b by means of a windowed sinusoidal signal sampled at 5 times the Nyquist frequency. The curvature objective J_L presents multiple peaks, dominated by the

discontinuity at the edges of the sinusoid. Considering the relative local variation in signal J_Δ the curvature objective can be smoothened yielding the blended heuristic J_C. More importantly, J_C no longer peaks at the discontinuities of the signal but also incorporates regions corresponding to the actual highest signal curvatures.

- **Error adaptive sampling - J_E**

The rationale to extend sampling adaptivity taking into account information about measurement uncertainty (represented as error bars) is to either gain a more accurate model in the region of high uncertainty by obtaining more exploratory points or to better characterise (reducing) the uncertainty by repeated sampling. Considering the signals of interest to be ergodic, the relative error in derived statistical quantities is pre-determined by the number of datapoints collected, the uncertainty level and associated level of signal fluctuation (c.f. Equation 2). A surrogate model is subsequently built based on relative error $\varepsilon_{\bar{x}}$ attributed to each measurement point utilising the procedure described in section 3.1.

- **Improvement adaptive sampling – J_I**

A third objective function mitigates the risk of over-exploiting regions of higher error or curvature without adding information. Inspired by the implicit assumption that a site x_{int} at which a great change in signal model $s^{(t)}(x_{int})$ has taken place since the previous sampling iteration $s^{(t-1)}(x_{int})$ is more likely to lie within an interval of interesting signal features, the improvement objective J_I quantifies this relative added information content at x_{int};

$$J_I = \frac{|\, s^{(t)}(x_{\text{int}}) - s^{(t-1)}(x_{\text{int}})\,|}{|\, s^{(t-1)}(x_{\text{int}})\,| + 1} \cdot \theta \quad \text{where} \quad \theta = \begin{cases} 1 & |\, s^{(t)}(x_{\text{int}}) - s^{(t-1)}(x_{\text{int}})\,| \leq \varepsilon_{\bar{x}} \\ 0 & else \end{cases} \qquad \textbf{(15)}$$

Parameter θ ensures that only significant changes to the surrogate model, exceeding the local measurement error $\varepsilon_{\bar{x}}$ are considered since such changes cannot be attributed to mere uncertainty. The improvement objective J_I thus favours regions where a demonstrated meaningful change has occurred during the latest sampling and discourages regions with low contribution to the overall picture.

- **Sample spacing adaptive sampling – J_h, J_{ho}**

To encourage sampling of the domain toward a space-filling scheme, sample separation is modelled as an RBF interpolant after Mackman and Allen (2010);

$$J_h(x_{\text{int}}) = \left(1 - \sum_{k=1}^{N_p} \gamma_k \psi(\| x_{k,s} - x_{\text{int}} \|) \right)_+ \qquad \textbf{(16)}$$

The term $(\cdot)_+$ signifies the cut-off function introduced in equation (6). Existing sampling sites $x_{k,s}$ have unity value enabling the determination of RBF coefficients γ_k. While the surrogate modelling concept applied is identical to the suggested RBF process with the exception of omitting the polynomial term, the use of the supremum nearest neighbour distance d_{snn} as support radius proved to yield more robust models compared to a constant R. This distance is the largest of the minimum distances between data sites;

$$R = \theta \cdot d_{snn} \quad \text{with} \quad d_{snn} = \underbrace{\sup}_{i=1\ldots N_p} \{ \underbrace{\min}_{\substack{k=1\ldots N_p \\ k \neq i}} (\| x_{i,s} - x_{k,s} \|) \} \qquad \textbf{(17)}$$

where θ is set to $1.25/4$. The spacing objective J_h is zero at existing sample locations and grows away from sample sites towards a value of unity as illustrated in Figure 4-a. Regions distanced from existing data sites are concomitantly favoured for exploration by J_h toward completely uniform space filling. In contrast the exclusion factor J_{ho} represents the limit of J_h with an infinitesimal support radius with zeros at the sample sites and unity elsewhere and serves to exclude any identical duplicate samples which may arise. Since the finest spacing of elements in J_{ho} is dictated by the spacing of the interrogation grid x_{int}, J_{ho} effectively fixes the finest spacing for newly proposed samples. This linkage is desirable since it reflects matrix conditioning considerations along with the available precision of the physical experiment.

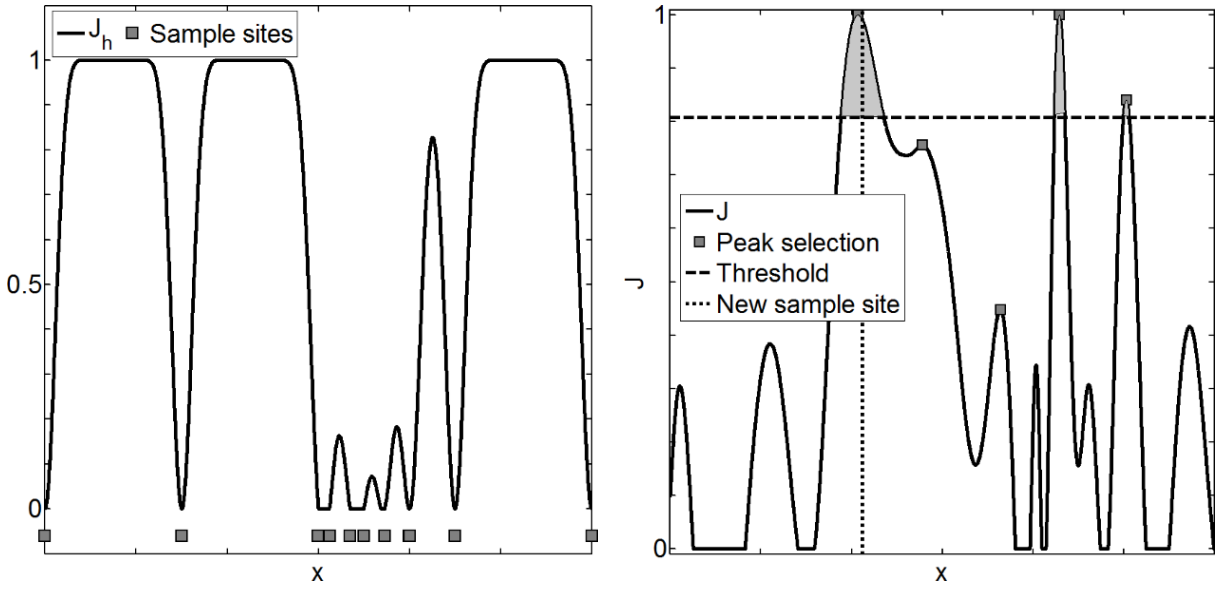


Figure 4: (a) The objective function J_h reaches zero at existing data sites and tends towards unity in between, imposing a tendency towards uniform space filling. (b) Illustration of sample selection utilising objective peak geometry; the new sample site is located at the centre of gravity (⋮) of the largest of grey areas highlighting sections of objective function exceeding the threshold (--). The threshold equals the mean of the 5 largest peak values (■).

- **Offset constants**

Offset constants are implemented to prevent their associated factors J_i from dominating the objective function's tendency towards zero in cases where $J_i \approx 0$. For the exclusion factor J_{ho}, which avoids redundant (coincident) samples, and J_h, which merely shapes the whole objective function according to sample spacing preference, this tendency to override the sampling objective function is desirable; a flat signal must still be sampled with equispaced samples. No offset factors are attributed to these object factors concomitantly. Domination of one of the other components (error, curvature or improvement) would undermine however the ability of adaptive sampling to target promising regions when just one of the sampling criteria is very small.

While the authors have not presented results, elaborate assessment has implied the following suitable range of offset values; $0.4 \leq \kappa_C \leq 0.6$, $0.4 \leq \kappa_E \leq 0.5$ and $0.4 \leq \kappa_I \leq 0.8$. Consequently a robust setting $\kappa_i = 0.5$ is proposed.

2.3 Sample selection

Each iteration a single new sample is prescribed. Although this process is computationally more intensive compared to the allocation of multiple new sites, the distribution of each unique sample is

ensured to be based on a maximum of information available about the response. Moreover, the selected process retains general applicability and reliability.

Because the objective function is the product of several different objective functions, smoothness is not guaranteed. Instead the resulting function may consist of multiple peaks attaining the maximum objective amplitude, distracting immediately neighbouring regions of nearly as high value. For this reason, rather than selecting the highest value in J as was done in prior adaptive methods, the new sampling location is decided based on objective peak geometry, i.e. the width and amplitude of all peaks exceeding a threshold value. Here the authors have adopted the principle that the integral of the difference between the objective function and threshold over an interval, relative to the integral of the entire domain acts as a heuristic for the allure held by that region.

The new sample selection routine is graphically depicted in Figure 4-b. A threshold is based on the mean value of at most the 5 highest peaks of J and is in this sense fully adaptive to the individual signal. The new sample is subsequently attributed to the interrogation grid point closest to the centroid of the dominant objective area rather than the peak location.

Relative to the domain width, the minimum sample separation still allowing for an invertible interpolation matrix is generally smaller than the finest spatial precision of the experimental equipment. This is certainly true of the Laser Doppler Anemometry (LDA) system used in the present work where the mechanical resolution of 0.01cm translates in 0.05% of the typical 10-20cm domain width. Accordingly, the discrete nature of the interrogation grid prohibits a degradation of the interpolation matrix's condition number.

2.4 Convergence Criteria for Automatic Stopping

While fully adaptive, the process thus presented does not come with criteria for deciding when to stop acquiring new samples. Two common stopping criteria are the predetermined budget of sampling points and an acceptable difference between the surrogate model and the underlying signal. In the former case this requires a-priori knowledge of both the spatial distribution of the signal and the adequate number of samples to represent such a function. Neither need be known to the experimentalist. The second condition is limited to use in benchmark testing of the sampling method rather than with real experiments because the true function needs to be known. The present work addresses these shortcomings, rendering the adaptive sampling method capable of complete automation. In accordance with experimental practice, the authors categorise the adaptive sampling process converged when the maximum in the improvement factor J_I, evaluated at the interrogation grid, falls below a threshold ε_{lim} for the last N_{CCI} iterations. Both threshold and number of consecutive iterations are updated adaptively.

The sampling objective function in this work includes two particularly salient emphases; regions of high uncertainty relative to response values and relatively dynamic response. The authors argue that if the measurement is contaminated by large errors, improvements must be ascertained to be related to the signal rather than an erroneous measurement. As such, a larger number of iterations is needed subjected to a more stringent threshold. Here, the non-dimensional error factor α is introduced, expressing the ratio between experimental uncertainty or error and amplitude range in underlying signal;

$$\alpha = \frac{\left|\overline{\varepsilon_x}\right|}{\left|s_{\max} - s_{\min}\right|} \qquad \textbf{(18)}$$

where $\left|\overline{\varepsilon_x}\right|$ is the mean of the absolute sample uncertainties and s_{max} and s_{min} are the maximum and minimum sample responses respectively. Recalling the definition of $\Delta s(x_{int})$ as per Equation 12 and

denoting the corresponding peak values by subscript p the signal contrast β, similar to image contrast, is evaluated through

$$\beta = \frac{\left|\Delta s_{p,\max} - \Delta s_{p,\min}\right|}{\left|\Delta s_{p,\max} + \Delta s_{p,\min}\right|} \qquad \textbf{(19)}$$

Values of contrast close to unity imply high amplitude features to dominate the objective function. A single dominant feature should not inhibit further exploration of the measurement domain however. For this reason a high signal contrast β must necessitate more sampling sites to capture such features and to encourage sampling of the wider domain. Finally, the error factor α and contrast β are combined in a linear scale for the automatic calculation of the convergence threshold (Figure 5);

$$\varepsilon_{\lim} = \varepsilon_{coarse} \cdot \left(1 - 0.9 \cdot \min\left(1, \tfrac{\alpha+\beta}{2}\right)\right) \qquad \textbf{(20)}$$

Parameter ε_{coarse} is a constant matched to the non-dimensional improvement factor J_I and relates to the level of repeatability in the data. Empirical studies performed by the authors have indicated values of ε_{coarse} ranging between $3 \cdot 10^{-3}$ for numerical studies (highly repeatable data) and 10^{-1} for experimental data (more erratic data) to yield robust stopping performances, independent of the response at hand. The required number of consecutive converged iterations N_{CCI} for which the maximum in J_I must fall below ε_{lim} is then expressed as

$$N_{CCI} = 1 + \text{ceil}\left[K_{CCI}\left(\frac{\alpha+\beta}{2}\right)\right] \qquad \textbf{(21)}$$

where K_{CCI} is a constant set to 10 to scale the number of required iterations in response to the average between contrast and error factor (Figure 5). Since N_{CCI} is determined adaptively for a given response rather than prescribed as a case-specific parameter, the method is completely automated. Note that now the necessity of more sampling sites ($^{\alpha+\beta}/_{2} \rightarrow 2$) is directly translated in an enhanced severity of the threshold ($\varepsilon_{lim} \rightarrow \varepsilon_{fine}$) and a concomitant larger number of required iterations satisfying the convergence criterion.

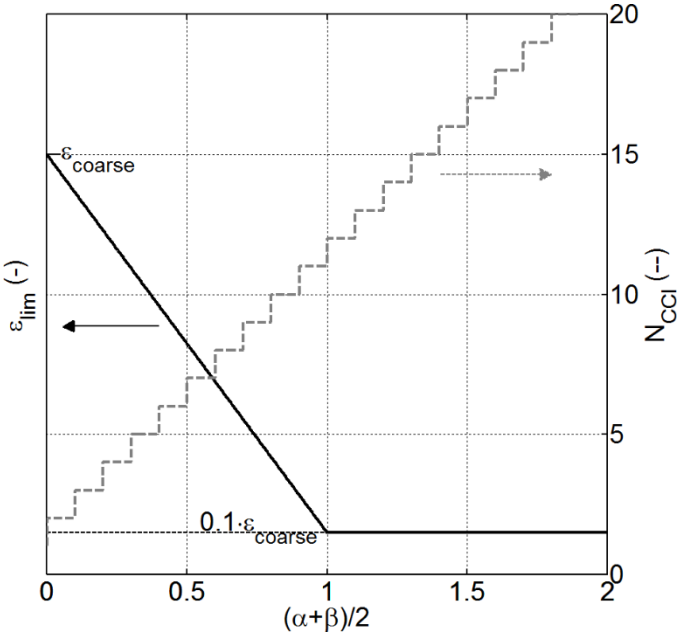


Figure 5: Evolution in convergence threshold ε_{lim} and number of converged iterations N_{CCI} with average between error factor α and contrast β.

2.5 Computational effort

The adaptive sampling approach requires several functions to predict new samples as indicated by the flowchart in Figure 2. While these processes may involve some computational overhead, the majority is related to the determination of the interpolation coefficients related to J_C, J_E and J_h. However, because of the use of Wendland functions the $(N_p+3)\times(N_p+3)$ matrix system defined in (5) is strictly positive definite. This allows the matrix inverse to be obtained by means of highly efficient Cholesky decomposition.

To demonstrate the limited additional computational cost, the ratio between the total time T and number of final samples N_p is presented in Figure 6 for the case of the flat plate boundary layer ($N_{p,0}=33$) to be discussed in section 5. A typical LDA datarate of 4kHz is assumed yielding a sampling time Δ_t of 1 second to gather 4000 instantaneous measurements at a fixed spatial location. Even though no efforts were made to further optimise the algorithmic implementation of the adaptive sampling process, Figure 6 indicates the experimental effort to be dominated by data collection ($T_{LDA}=N_p\cdot\Delta_t$). The pie chart reveals the majority of the overhead (T_{CPU}) to be attributed to the construction of the surrogate model and determination of the objective factors J_C and J_E. Selection of new samples based on the objective function J covers 6% of the computational effort compared to 10% to establish J_h.

Figure 6 shows that the adaptive sampling scheme demands a marginal time increase. However, the adaptive sampling strategy allows half of the equidistant sampling points to be omitted as will be shown in the remainder. The authors found that the resulting overall time-gain outweighs any higher computational cost compared to traditional sampling as optimised sampling resulted in a typical speed up of data collection by a factor ranging between 2 and 3. It should be noted that even higher time gains can be obtained as the datarate decreases (larger Δ_t).

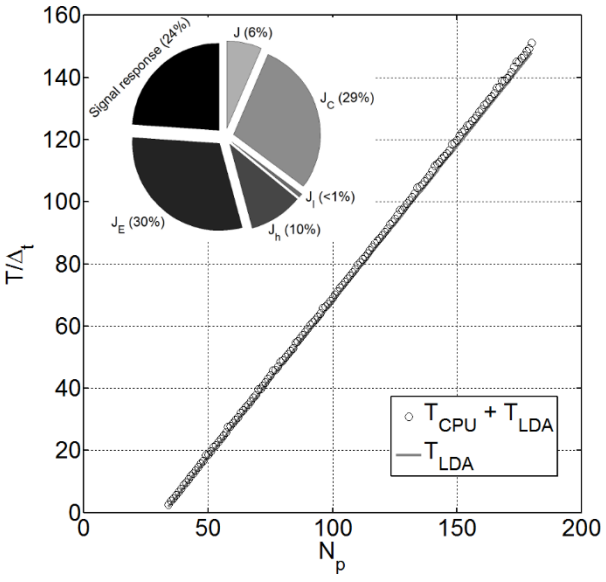


Figure 6: Comparison between the data collection time (T_{LDA}) and added computational time for adaptive sampling (T_{CPU}) as a function number of samples and normalised with the measurement time for 1 spatial location (Δ_t). Data is based on the experimental boundary layer measurement initiated with 33 samples ($N_{p,0}=33$). The inset shows the distribution of T_{CPU} across the different processes within the proposed adaptive sampling strategy.

3. Numerical Experiment

The versatility and robustness of the new adaptive sampling method is first demonstrated using synthetic experiments. Underlying signals and spatial distributions in error are mathematically defined

and chosen as to pose important challenges to the scheme. In all test cases, a small initial grid of uniformly spaced samples is provided and the adaptive method proceeds to add one new sample per iteration until the automatic convergence criterion is satisfied. In each test case the surrogate model obtained with adaptive sampling is assessed against that obtained from uniform (full factorial) sampling.

The improved potential to represent the underlying function *f* is quantified by means of the traditional root-mean-square error (RMSE). To obtain unbiased estimates a cubic interpolation is applied between the samples yielding the response *g(x)*.

$$RMSE = \sqrt{\frac{1}{N_e}\sum_{i=1}^{N_e}(g(x_i) - f(x_i))^2} \qquad \textbf{(22)}$$

However, the RMSE is only a global metric of modelling error and does not properly assess performance where experimental uncertainty plays a role; it ignores the fact that a model may have mediocre performance in mundane regions of a response while capturing the most important response features. In preserving accuracy of the model representing the true function the adaptive sampling method aims to place samples in regions of interest, notably zones of high curvature and high uncertainty. To this extent two additional performance quality measures, Q_{JE} and $Q_{J\Delta}$, based on the objective functions for experimental uncertainty (J_E) and curvature (J_Δ) are introduced defined as

$$Q_X = \frac{\varsigma_{uniform}}{\varsigma_{adaptive}} \quad \text{where} \quad \varsigma = \frac{1}{N_p}\int X \cdot \delta(x - x_i)\,dx = \frac{1}{N_p}\sum_{i=1}^{N_p} X(x_i) \quad \text{and} \quad X=J_E \text{ or } J_\Delta \qquad \textbf{(23)}$$

Here $\delta(x-x_i)$ symbolises the final distribution of N_p sampling sites x_i. In this form, the larger Q_X, the more correlated the sample distribution is with the target objective function component and the higher its quality according to the criterion. It should be noted that whilst the quality metrics cannot be compared between different test cases, they allow comparison of different surrogate models or methods within the same test case.

- **Bi-modal non-homogeneous signal**

The first experiment considers the case of two Gaussian functions with elevated error near the centre of the domain. The analytical description of the 1D signal, its Laplacian and error are given as

$$\begin{aligned} f(x) &= 0.1\cdot\exp(-150\cdot(\tfrac{x}{L}-\tfrac{1}{3})^2)+1\cdot\exp(-1500(\tfrac{x}{L}-\tfrac{5}{6})^2) \\ \nabla^2 f(x) &= 30\cdot\exp(-150\cdot(\tfrac{x}{L}-\tfrac{1}{3})^2)\cdot(300\cdot(\tfrac{x}{L}-\tfrac{1}{3})^2-1)+300\cdot\exp(-1500(\tfrac{x}{L}-\tfrac{1}{3})^2)\cdot(3000(\tfrac{x}{L}-\tfrac{1}{3})^2-1) \\ \varepsilon_x(x) &= 0.1\cdot(\exp(-750\cdot(\tfrac{x}{L}-\tfrac{3}{5})^2)+1) \end{aligned} \qquad \textbf{(24)}$$

This synthetic test case is devised to demonstrate the impact of measurement error. The domain length was set to *L=1.5*. Starting with 9 uniformly spaced samples the iterative evolution in signal response is depicted in Figure 7-a. Adaptive samples are gradually concentrated on peaks in the Laplacian and regions of larger uncertainty even though the surrogate model value in this area is nearly flat and would garner little interest based on curvature alone. This is further illustrated in Figure 7-b by interpreting $\nabla^2 f(x)$ and ε_x as probability density functions (effectively J_L and J_E respectively) and comparing their combined cumulative density function (cdf) with the sample distributions for the standard full factorial and adaptive sampling method. The adaptive approach imposes a total of 48 sampling sites after 40 (=9+1×39) iterations. To facilitate unbiased comparison between results, the same budget of samples was used in the traditional method. Here samples are distributed equidistantly irrespective of the underlying signal, yielding a linear cdf. The adaptive approach on the other hand yields a distribution resembling the combined cdf as imposed by the product $J_L \cdot J_E$. The resulting cdf

can of course only approximate theory since the latter does not consider additional criteria such as space filling (J_h) and improvement (J_I) or smoothing operations (*MA*).

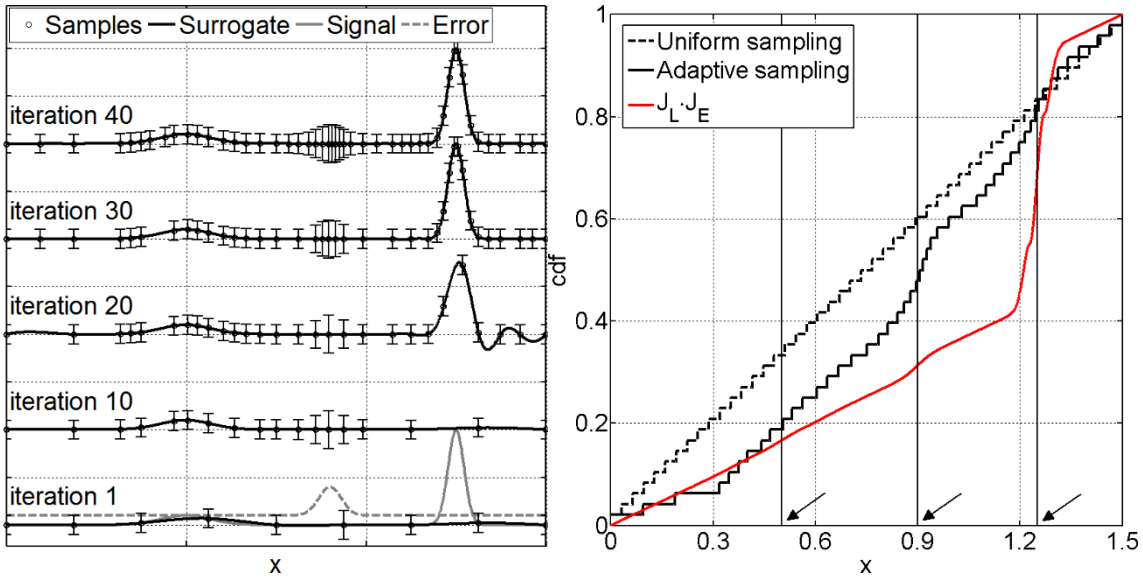


Figure 7: (a) Evolution in adaptive surrogate model with iteration number for the case defined by (23) and $N_{p,0}=9$. The imposed signal and error are depicted in solid and dashed gray lines respectively. **(b)** Cumulative density function of the sampling distribution as per uniform sampling (--), adaptive sampling (−) and theory (−). Arrows indicate the centre of the Gaussians defining the signal (outer) and error (middle).

As a result of the ameliorated sampling, despite utilising an equal amount of total number of samples N_p, the adaptive method can yield a drastic reduction in root mean square error (RMSE) compared to the full factorial method (Table 1). In fact, the quality metric for curvature confirms a much stronger capture of high-curvature regions by the adaptive sampling scheme than by full factorial sampling as is evidenced by the distribution of sampling points (Figure 7-b). Simultaneously, the adaptive process concentrates more samples in the region of larger error as quantified by Q_{JE}. With increasing initial sampling sites, the number of new measurement locations decreases and adaptive sampling becomes less advantageous. For the present test case, beyond $N_{p,0}=65$ all signal features are sampled sufficiently as seen by $Q_{J\Delta}$ tending towards unity, necessitating no further characterisation. The imposed convergence criteria correctly self-terminate the recursive procedure after 7 iterations.

Table 1: Performance comparison between uniform and adaptive sampling in case of a double Gaussian with varying error distribution as per (23).

	Test case: double Gaussian with inhomogeneous error distribution						
$\mathbf{N_{p,0}}$	5	9	17	33	65	129	257
$\mathbf{N_p}$	51	48	47	68	74	136	267
$\mathbf{RMSE_{Uniform}}$	0.0176	0.128	0.0215	0.0058	0.0047	0.0022	0.0002
$\mathbf{RMSE_{Adaptive}}$	0.0022	0.0027	0.0029	0.0006	0.0025	0.0005	0.0001
$\mathbf{Q_{J\Delta}}$	1.9728	1.9363	2.3437	2.6893	2.0292	1.3228	1.1791
$\mathbf{Q_{JE}}$	2.7547	2.7982	2.7733	2.8572	2.7980	2.7143	2.6752

Figure 8 shows the convergence history for the test case considered, illustrating how the convergence criteria adapt as more is learned about the response. The discovery of a new feature at iteration 15 leads to a substantial increase in J_I exceeding the threshold ε_{lim}. The new feature corresponds to the Gaussian with the larger amplitude, causing both non-dimensional error α and contrast β to decrease.

This subsequently increases the threshold and reduces the required number of consecutive converged iterations. As all prominent features are becoming appropriately sampled, the maximum improvement decreases and remains below ε_{lim} causing the recursive process to stop.

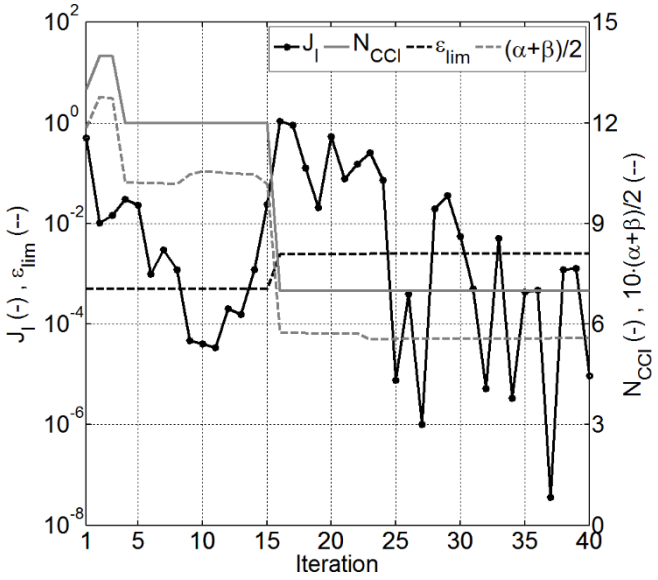


Figure 8: Evolution in convergence threshold ε_{lim} and number of converged iterations N_{CCI} with average between error factor α and contrast β.

Wiggles can be observed in the response model on the right hand region of the highest Gaussian peak in iteration 20. These concomitantly introduce local regions of high artificial curvature and are attributed a sampling site in the subsequent iteration causing the oscillations to disappear. Such oscillations are equivalent to the well-known Runge phenomenon (Runge, 1901) inherent to polynomial interpolation. In the framework of adaptive sampling such wiggles are beneficial. First, they stimulate the excursion of the entire domain. The second reason relates to interpolation; the wiggles indirectly identify sampling locations important in the reconstruction of the measured signal. By its very nature the appearance of such wiggles is thus minimised when interpolating the data in the post-processing stage. This is further exemplified in the second test case.

Although the example presented demonstrates the proposed sampling method to achieve promising results, the technique will not remedy poor experimental practice in that the performance of the adaptive sampling strategy, in case of strongly inhomogeneous information content, does depend on the initial sampling distribution. As an example, considering an isolated peak on an unchanging background, the adaptive method cannot home in on the peak unless information is available indicating the presence of this feature. Initial measurements must thus provide at least an indication of features present for the adaptive routine to determine the signal could be of interest in that region.

- **Linear chip**

In the second example the adaptive routine is applied to a linear chirp signal of amplitude 1 with an initial frequency of unity and chirp rate of 20; $f(x)=1\cdot\sin(2\pi\cdot(x+10\cdot x^2))$ with $0\leq x\leq 1.5$. The frequency thus varied spatially following $1+20\cdot x$. A constant error ε_x was imposed. The evolution in surrogate model with iteration is depicted in Figure 9-a starting with 17 initial samples $N_{p,0}$, well below the Nyquist frequency. In this case the iterative routine is stopped after 152 recursive iterations equating to a total of 167 sampling points. In comparison, the optimal number of samples is calculated by imposing N_k samples between each of the zero crossings. While each semi-wave is thus sampled with equidistant points, the spacing varies spatially in a discontinuous manner. The total number of

samples over the considered domain is then given by the linear equation $48 \cdot (N_k - 1)$. Given the 167 samples, this equates in average to 9 sites per oscillation, satisfying the Nyquist criterion. Alternatively, retaining the spatial resolution at the smallest wavelength a uniform sampling approach would demand 278 sites, nearly double that of the adaptive routine.

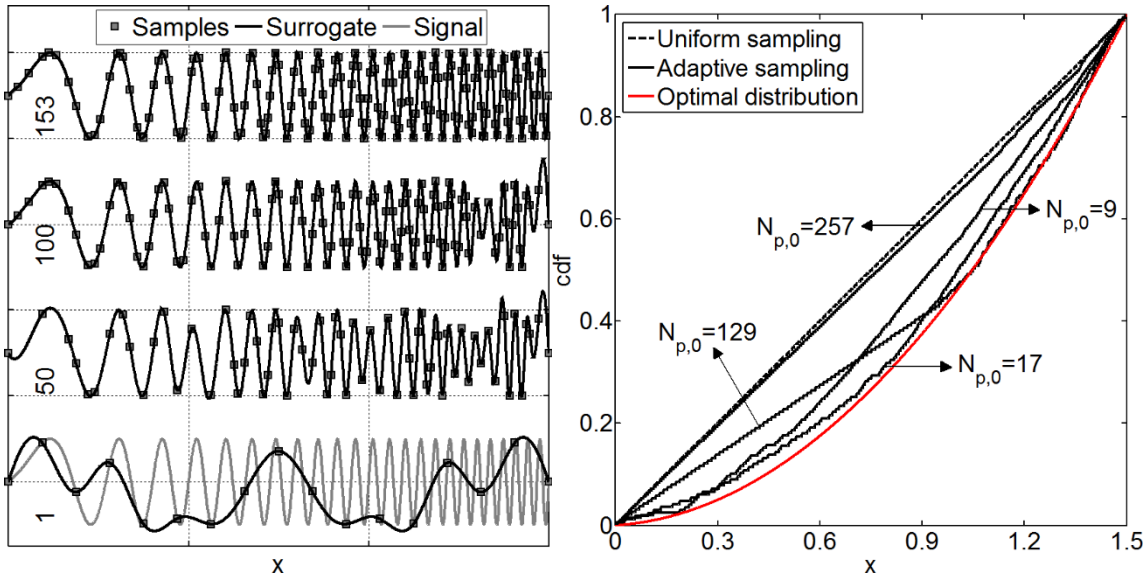


Figure 9: (a) Evolution in adaptive surrogate model with iteration number for linear chirp with chirp rate 20 and $N_{p,0}=17$. The underlying signal is depicted in solid gray while a constant background error was imposed. **(b)** Cumulative density function of the sampling distribution as per uniform sampling (--), adaptive sampling (−) and theory (−) for varying number of equally spaced initials samples.

Table 2: Performance comparison between uniform and adaptive sampling in case of a linear chirp signal for varying initial sampling density.

Test case: linear chirp signal with chirp rate 20						
$\mathbf{N_{p,0}}$	9	17	33	65	129	257
$\mathbf{N_p}$	198	167	152	204	221	268
$\mathbf{RMSE_{Uniform}}$	0.0388	0.0522	0.0747	0.0359	0.0299	0.0201
$\mathbf{RMSE_{Adaptive}}$	0.0353	0.0386	0.0561	0.0252	0.0241	0.0199
$\mathbf{Q_{J\Delta}}$	1.2439	1.3974	1.2515	1.3273	1.2711	1.001

Considering the spatial distribution of the frequency content, the cumulative probability of the ideal sampling distribution is given by its integral i.e. $cdf(x)=x+10 \cdot x^2$. Applying an equal number of equispaced samples on the other hand ignores this spatially varying information, yielding a linearly evolving cdf (Figure 9-b). When gradually increasing the initial number of data sites for the adaptive routine, the cumulative density function of the final sample distribution approaches the ideal distribution. Resultantly the quality factor $Q_{J\Delta}$ increases (Table 2), indicative of the majority of samples being concentrated near the signal peaks. In turn the adaptive approach also yields an *RMSE* inferior to the standard technique. Contrary to the full factorial sampling, data post-processing may require the non-equidistant distributed data to be re-interpolated onto a structured grid. The demonstrated adherence of the adaptive sampling to the signal's frequency content implies a minimisation of potential interpolation errors as indicated by the relevant RMSE values.

Increasing the number of initial points again proves not to be necessarily beneficial. As $N_{p,0}$ augments, all frequencies of the chirp signal gradually become properly sampled, necessitating increasingly fewer new adaptive sites and moving the adaptive cdf towards a linear tendency (Figure 9-b). In the cdf for $N_{p,0}=129$ this onset to a linear tendency for the lower frequency part of the signal is clearly

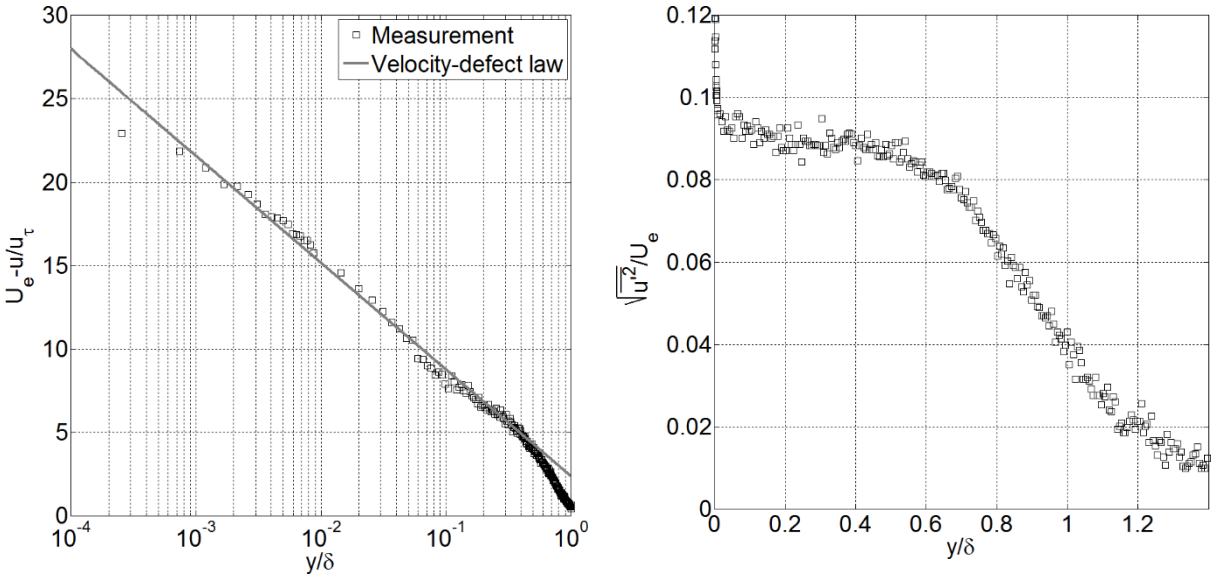
visible. The adaptive method thus offers little advantage when a very dense initial sampling is utilised; a signal with homogeneous information content is known to be sampled optimally by a uniform scheme satisfying the Nyquist sampling criterion. Indeed, both *RMSE* and $Q_{J\Delta}$ imply both full factorial and adaptive sampling to have nearly identical performances at high $N_{p,0}$.
Arguably the above implies traditional equidistant sampling to be conducive. However, without a-priori knowledge of the signal, an experimentalist is forced to update the required number of samples on the basis of previous results until satisfactory. Either this results in highly inefficient data collection or local refinements in data spacing. Conversely, the adaptive routine is fully autonomous, placing samples optimally, requiring no or a-priori knowledge of the signals and no user interaction. The adaptive routine thus reflects within a mathematical framework the decision making process of the experimentalist.

4. Experimental Application

The proposed adaptive method suggests new sample locations iteratively and is therefore best suited to experiments in which data are collected serially. To demonstrate the potential of the proposed method, both adaptive sampling and standard (uniform) sampling was applied to single component LDA-based measurements. Experiments were conducted in the University of Bristol's Low Turbulence Wind Tunnel. This tunnel has an octagonal test section (0.8 m×0.6 m) and attains turbulence levels below 0.05%. Wake surveys were performed by means of a two-component Dantec Dynamics Laser Doppler Anemometry system operating in crossed beam mode. Prior to passing the laser beams through optics with focal lengths of 600 mm, an expansion of factor 1.98 was applied yielding a measurement volume extending 0.17mm in streamwise normal direction. At each measurement location velocity statistics were evaluated on the basis of typically 4000 instantaneous samples ($N_s \approx 4000$).

- **Artificially thickened flat plate boundary layer**

The first experimental test case investigated the velocity profile across an artificially thickened boundary layer on a flat plate with sharp leading edge. The trip mechanism consisted of a 4mm wire placed 10cm downstream of the leading edge. Benchmark boundary layer characteristics were obtained from 427 LDA measurements adopting two different sampling schemes; in the near vicinity of the wall ($y^+<57.9$) measurement locations were separated 0.05mm equivalent to 1.9 wall-units, whereas the partition extending into the freestream ($y^+\approx 8.5\cdot 10^3$) was sampled at $\Delta y^+=22.7$. With a local freestream velocity of $U_e=25.38 ms^{-1}$ 20cm downstream ($Re_x=0.167\cdot 10^6$) the trip wire produced a turbulent boundary layer with shape factor $H=1.35$ and ratio between boundary layer thickness (99% freestream velocity) and displacement thickness of $\delta/\delta^*=7.5$. Very close to the wall reliable velocity results were hampered by the finite extent of the LDA measurement domain (7 wall-units) and low data rate. The wall-friction velocity was therefore estimated by fitting the velocity-defect law, $(U_e-u)/u_\tau=-\kappa^{-1}\cdot ln(y/\delta)+B$, in the region $10^{-2.9}\leq y/\delta\leq 0.4$ yielding $\kappa=0.36$, $B=2.4$ and $u_\tau=0.61 ms^{-1}$ (Figure 10-a). Such characteristics are in close agreement with those reported by Klebanoff and Diehl (1952) and Rona and Soueid (2010). The positive value of constant B indicates the presence of a mild adverse pressure gradient, which is also advocated by the shape in longitudinal turbulence intensity in Figure 10-b (Schloemer, 1966).



(a) **(b)**

Figure 10: Artificially thickened flat plate boundary layer, 20cm downstream of a 4mm trip ($Re_x=0.167\cdot10^6$) **(a)** Comparison of experimental results at fine resolution with velocity-defect law yielding friction velocity $u_\tau=0.61ms^{-1}$. **(b)** Longitudinal turbulence intensity.

The adaptive sampling routine was initiated with 17 samples ($N_{p,0}=17$). Because it was assumed no a-priori knowledge of the exact wall location is available, the measurement domain extended 2mm below the surface ($-0.02\leq y/\delta\leq2.1$). In total 99 samples were placed adaptively before the process automatically stopped. Contrary to the sampling distribution obtained by standard equidistant sampling when using an equal number of data sites, adaptivity favours the region close to the wall and adopts sparse sampling towards the freestream. The corresponding probability closely follows the curvature in the velocity profile and turbulence intensity as illustrated in Figure 11. The sample nearest to the wall as per adaptive sampling was located at $y^+=6$ compared to $y^+=75$ with the standard routine, further enabling a better capture of the turbulence peak and viscous sub-layer.

Because of scatter in velocity and RMS data, doubling the number of initial points resulted in an increased number of final data points ($N_p=142$). The probability distributions of samples remained unchanged however with respect to the $N_{p,0}=17$ case. Because of the higher number of samples the standard approach now placed the first sample at $y^+=22$ whereas the new method maintained the first sample at $y^+=6$, manifesting its robustness. When continuously providing a better initial characterisation of the boundary layer by further increasing $N_{p,0}$, the adaptive process consistently stopped after placing a total of approximately 240 (±30) samples. Here the sampling process was eventually driven by the oscillations in velocity data due to measurement uncertainty (Figure 11). However, the sampling distribution retained the unimodal tendency in the region of higher velocity curvature and uncertainty yielding a better resolved velocity profile in the near-wall region substantiated by a y^+ of 2.7 compared to a $y^+=5$ with equidistant sampling. The latter highlights the advantage of the proposed sampling method as uniform sampling required nearly double the number of measurement locations to attain comparable spatial resolution.

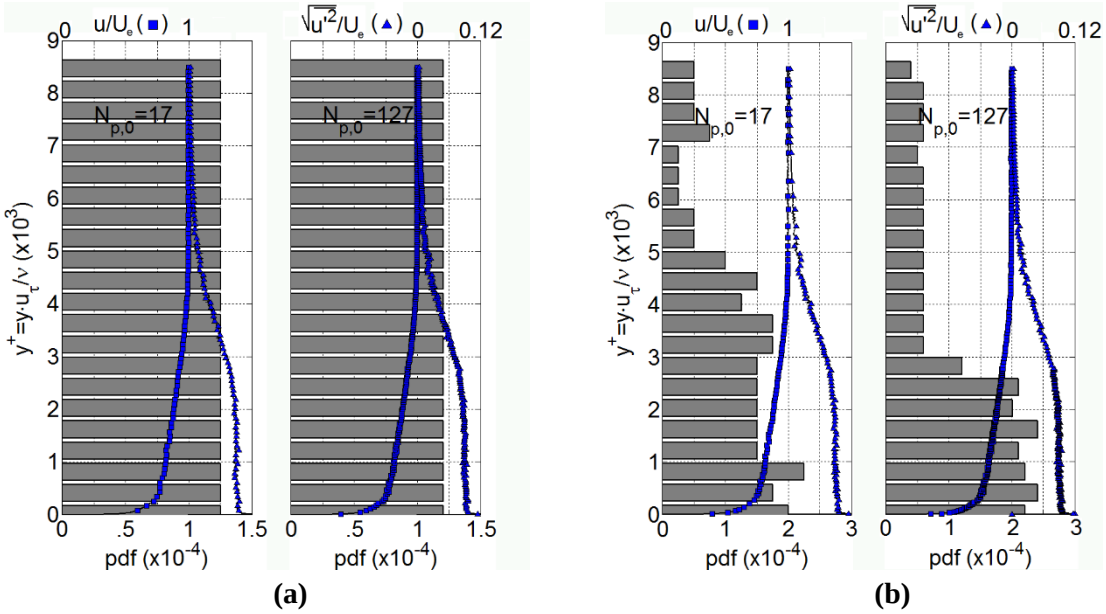


Figure 11: Probability density function (pdf) in sample distribution (N_p=*99*) when initiating the adaptive process with 17 and 127 samples; **(a)** uniform sampling **(b)** adaptive sampling. Measured velocity (■) and turbulence intensity profiles (▲) are superimposed for clarity.

- **NACA0012 near wake**

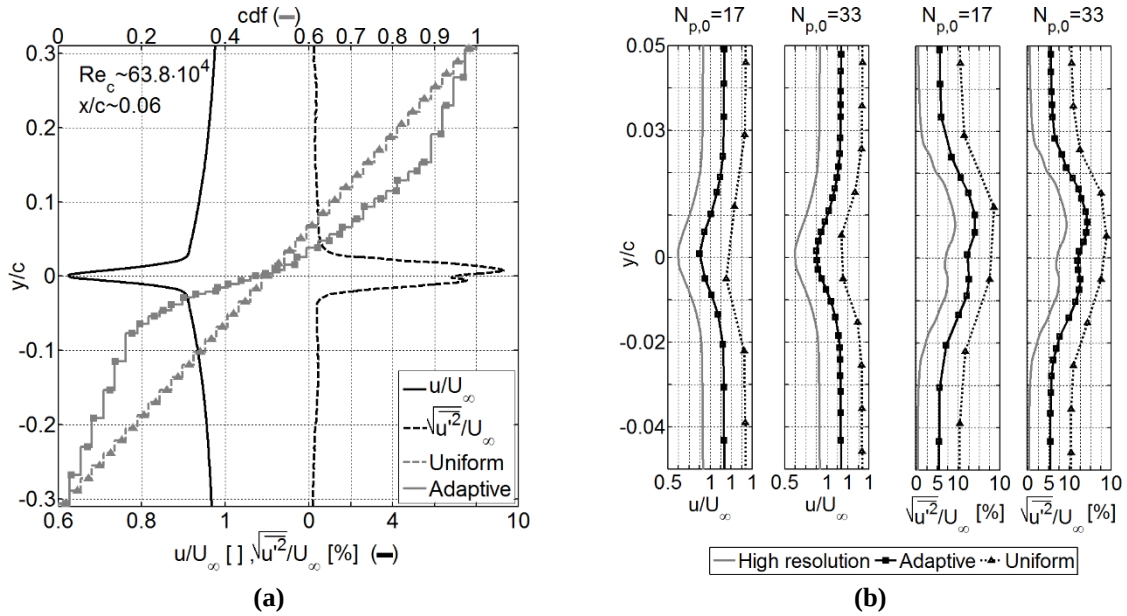
In the second experimental application a wake survey behind a NACA0012 airfoil placed 0º incidence at a chord-based Reynolds number of $Re_c=63.8\cdot10^4$ was performed. The airfoil had a slot on the suction side at 70% chord length to facilitate studies related to boundary layer suction and blowing. Such concepts are not represented in the current study though. The measurement station was positioned 2cm downstream of the trailing edge ($x/c\approx0.065$) while the considered domain extended approximately 0.65 chord lengths perpendicular to the freestream ($-0.32\leq y/c\leq0.32$). A wake survey at high resolution ($\Delta y/c\sim0.003$) was performed to facilitate the assessment between the standard and advanced sampling strategy. Resulting horizontal velocity and turbulence intensity profiles are depicted in Figure 12-a. A velocity deficit of 40% can be observed centered at $y/c=0$. Figure 12-a further indicates the domain extent to be insufficient to reach freestream conditions, which according to Bairstow (1946) should take place for y/c exceeding 0.05 in absolute value. As noted by Cooper (1984) in the case of a flat plate, close to the trailing edge the wake is not yet developed. Instead upper and lower boundary layers can be seen to merge in the horizontal velocity fluctuations additionally showing the presence of a double peak. Here the observed asymmetry in turbulence intensity is caused by early separation of the boundary layer on the upper side of the airfoil due to the presence of the slot. These findings are in good agreement with the data of Kim *et al.* (2009) for a clean NACA0012 airfoil at 3º incidence.

The adaptive sampling process was initiated with 17 initial sampling points and self-terminated after 20 iterations with $\varepsilon_{lim}=0.1$ as per the authors' suggestion, yielding a total of 37 sampling points. The distribution across the domain in terms of the cumulative density function (cdf) is overlaid on the signal in Figure 12-a. A sharp increase in the cdf can be observed in the region of larger turbulence intensity and velocity deficit, contrary to uniform sampling (adopting an equal number of equispaced samples) which follows the typical linear tendency. Pivotal is the outcome of the proposed routine when interpolating the data using a cubic scheme as displayed in Figure 12-b. Because of the innate sampling concentration in the wake, the adaptive routine is capable of capturing both the velocity minimum and bi-modal shape in the turbulence intensity as opposed to the standard methodology. Based on the sampling density noticeable in Figure 12-b, standard uniform sampling would require at

least twice as many measurement stations as adaptive sampling to achieve the same spatial resolution. When doubling the number of initial measurements ($N_{p,0}=33$), the automatic scheme places 28 additional sites concentrated in the wake as illustrated in Figure 12-b, thus detailing every feature while the outer region remains sparse. This is further quantified by evaluating the drag coefficient C_d on the basis of momentum exchange;

$$C_d \approx 2\int_{-\infty}^{+\infty} {}^{u}\!/_{U_\infty}\left(1-{}^{u}\!/_{U_\infty}\right)d\,{}^{y}\!/_{c} \qquad \textbf{(25)}$$

Both the high resolution and adaptive technique yield a C_d of 0.0439 independent of $N_{p,0}$. This exemplifies the potential of the new routine to optimise the distribution of measurement locations. The under-resolved full factorial data returns a value 4 drag counts higher, i.e. 0.0443 in case of $N_{p,0}=17$ and only tends towards the benchmark value with increasing initial sampling (0.0441 for $N_{p,0}=33$). Still, Figure 12-b verifies that the turbulence intensity profile is not well represented. It should be noted that because freestream conditions ($u/U_\infty=1$) are not reached in the present measurement domain ($-0.32\leq y/c\leq 0.32$), evaluation of the above integral will yield incorrect values for drag coefficients and can only be used for juxtaposition of the sampling strategies.



(a) **(b)**

Figure 12: Wake survey behind a NACA0012 aerofoil of 31cm chord length, 2cm downstream of the trailing edge at $Re_c\approx 63.8\cdot 10^4$ **(a)** Horizontal wake velocity and turbulence intensity profile superimposed with the spatial distribution of sampling points ($N_{p,0}=17$) **(b)** Close-up of the comparison between the sampling strategies in terms of velocity deficit and streamwise turbulence intensity.

5. Conclusions

Without *a-priori* knowledge, the most common sampling approach is to place equidistant samples along the parameter space describing the underlying process. Here measurement locations are distributed uniformly irrespective of the underlying signal. Sampling locations should however concentrate in regions of higher curvature and measurement uncertainty. While the former is shown to reduce further interpolation errors, the latter sampling criterion will enhance data reliability. To this extent a novel sampling methodology has been presented in this paper, which incorporates mathematical formulations of the typical decisions taken by an experimentalist. This includes a criterion to automatically stop the measurement when sufficient data is obtained, minimising the need for user-input.

The “smart” sampling method iteratively reconstructs a surrogate model of the underlying signal by means of Radial Basis Functions. Regions of higher data curvature and data uncertainty are subsequently attributed a higher number of sampling stations, though balanced by sample separation as to avoid clustering. Demonstrated improvement in the surrogate model serves as a final heuristic to dictate successive data extraction sites. These criteria are combined in a unique objective function on the basis of which new sampling sites are selected. Stopping criteria combine local alterations in surrogate model as a result of adding new information and local uncertainty to quantify convergence of the recursive process.

The potential and robustness of the new sampling methodology has been substantiated by applications to computer generated signals and experimental Laser Doppler Anemometry measurements. For signals with inhomogeneous information content, the adaptive method provides sampling schemes which capture signal features more accurately and/or with fewer samples compared with uniform sampling. Alternatively, for signals with homogenous information content, for which uniform sampling is the optimal design, the adaptive method recovers toward this result.

The novel adaptive routine is shown to allow for fully autonomous, highly efficient data collection, requiring no a-priori knowledge of the signals and no user interaction. Moreover, the adaptive routine generally required half the number of samples compared to traditional methods while retaining spatial resolution.

The presented concepts have been shown to be preferential in one-dimensional experiments but can be extended into higher dimensions. Future work will therefore focus on incorporating the suggested adaptive sampling method into PIV image processing.

References

Bairstow L (1946) Applied Aerodynamics. Longmans Green

Bendat J S, Piersol A G (1966) Measurement and analysis of random data. Wiley, New York

Chow S C (2014) Adaptive Clinical trial design. Annual Revision of Medicine 65: 405-415

Chow S C, Chang M (2008) Adaptive design methods in clinical trials – a review. Orphanet Journal of Rare Diseases 3 (11). doi: 10.1186/1750-1172-3-11: pp.

Congdon C D, Martin J D (2007) On using standard residuals as a metric of kriging model quality, AIAA 2007-1928, 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Honolulu, Hawai

Dassi F, Mola A, Si H (2014) Curvature-adapted remeshing of CAD surfaces. Procedia Engineering 82: 253-265

Fasshauer G E (2007) Meshfree Approximation Methods with Matlab. World Scientific, NJ.

Fornberg B, Driscoll T, Wright G, Charles R (2002) Observations on the behavior of Radial Basis Function approximations near boundaries. Comput Math Appl 43: 473-490.

Forrester A I, Keane A J (2009) Recent advances in surrogate-based optimization. Prog Aero Sci 45: 50-79

Forrester A I, Keane A J, Bressloff N W (2006) Design and analysis of ”noisy” computer experiments. AIAA Journal 44(10): 2331–2339

Forrester A, Sóbester A, Keane A (2008) Engineering design via surrogate modelling – a practical guide. John Wiley & Sons Ltd., UK

Fox J (1997) Applied regression analysis, linear models, and related methods. Sage Publications, Beverley Hills, CA.

Hardy R L (1971) Multiquadric equations of topography and other irregular surfaces. J Geoph Res, 76: 1905-1915

Hussain F, Burton R R, Joshi S B (2002) Metamodeling: radial basis functions versus polynomials. European Journal of Operational Research, 138:142–154

Hussain M F, Burton R R, Joshi S B (2002) Metamodeling: Radial basis functions, versus polynomials. Eur J Oper Res 138: 142-154.

Ilzarbe L, Álvarez M J, Viles E, Tanco M (2008) Practical applications of design of experiments in the field of engineering: a bibliographical review. Quality and Reliability Engineering International, 24: 417-428

Kim D-H, Yang J-H, Chang J-W, Chung J (2009) Boundary layer and near-wake measurements of NACA 0012 airfoil at low Reynolds numbers, AIAA 2009-1472,47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition, Orlando, Florida

Klebanoff PS, Diehl Z W (1952) Some features of artificially thickened fully developed turbulent boundary layers with zero pressure gradient, naca-report-1110

Krishnamurthy T (2005) Comparison of response surface construction methods for derivative estimation using moving least squares, kriging and radial basis functions. AIAA 2005-1821, 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Austin, Texas

Krishnamurthy T (2003) Response surface approximation with augmented and compactly supported radial basis functions, AIAA 2003-1748, 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference, Austin, Texas

Larsson E, Fornberg B (2005) Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. Computers and Mathematics with Applications 49: 103–130

Lovison A, Rigoni E (2010) Adaptive sampling with a Lipschitz criterion for accurate metamodeling, Communications in Applied and Industrial Mathematics 1(2): 110-126.

Mackman T J, Allen C B (2010) Investigation of an adaptive sampling method for data interpolation using radial basis functions. International journal for numerical methods in engineering 83: 915-938

Mackman T, Allen C, Ghoreyshi M, Badcock K (2013) Comparison of adaptive sampling methods for generation of surrogate aerodynamic models. AIAA Journal 51 (4): 797–808

McKay M D, Conover W J, Beckman J R (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2): 239-245

Paiva R M, Carvalho A R D, Crawfordy C, Sulemanz A (2009) A Comparison of surrogate models in the framework of an MDO tool for wing design. AIAA 2009-2203. 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Palm Springs, California.

Picheny V, Ginsbourger D, Roustant O, Haftka R T, Kim N-H (2010) Adaptive designs of experiments for accurate approximation of a target region. Journal of Mechanical Design 132 (7). doi:10.1115/1.4001873

Rippa S (1999) An algorithm for selecting a good value for the parameter c in radial basis function interpolation. Adv Comput Math 11: 193-210

Rona A, Soueid H (2010) Boundary layer trips for low Reynolds number wind tunnel tests, AIAA 2010-399, 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Conference Proceedings, Orlando Florida

Runge C (1901) Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. Zeitschrift für Mathematik und Physik 46: 224–243

Sacks J, Welch W J, Mitchell T J, Wynn H P (1989) Design and analysis of computer experiments. Statistical Science 4(4): 409–435

Scheuerer M (2011) An alternative procedure for selecting a good value for the parameter c in RBF-interpolation. Adv Comput Math 34: 105-126

Schloemer H H (1967) Effects of Pressure gradients on turbulent-boundary-layer wall-pressure fluctuations. The journal of the acoustical society of America 42(1): 93-113

Shannon C E (1949) Communication in the presence of noise. Proc. Institute of Radio Engineers, 37: 10–21

Sparta M, Toffoli D, Christiansen O (2009) An adaptive density-guided approach for the generation of potential energy surfaces of polyatomic molecules. Theoretical Chemistry Accounts 123: 413–429

Theunissen R, Di Sante A, Riethmuller M L, Van den Braembussche R A (2008) Confidence estimation using dependent circular block bootstrapping: application to the statistical analysis of PIV measurements. Exp Fluids 44: 591-596

Theunissen R, Scarano F, Riethmuller M L (2007) An adaptive sampling and windowing interrogation method in PIV. Meas Sci Technol. 18: 275-287.

Toal D J J, Bressloff N W, Keane A J (2008) Kriging hyperparameter tuning strategies. AIAA Journal 46 (5): 1240-1252

Weissman S A, Anderson N G (2014) Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. Organic Process Research & Development, doi:10.1021/op500169m

Wendland H (2005) Scattered Data Approximation. Cambridge University Press, Cambridge, U.K.

Yu M, Zhang Y, Li Y, Zhang D (2013) Adaptive sampling method for inspection planning on CMM for free-form surfaces. International Journal of Advanced Manufacturing Technology 67:1967–1975